1 Frame Theory

If $\operatorname{rank}(A) = m$ for $A \in \mathbb{C}^{n\times m}, n \geq m$, then $A^\dagger = (A^*A)^{-1}A^*$ is a left-inverse of A and the solutions to $LA = I$ are

$$L = A^\dagger + M(I - AA^\dagger), \quad M \in \mathbb{C}^{n\times m}.$$

Definition 1.1 (Frame)**.** *Let H be a Hilbert space. A set $(g_k)_k \in H$ is called a frame if there exist $0 < A \leq B < \infty$ such that*

$$A\|x\|^2 \leq \sum_k \|\langle x, g_k\rangle\|^2 \leq B\|x\|^2, \quad \textit{for all } x \in H.$$

Definition 1.2 (Analysis & Synthesis)**.**
$T\colon H \to \ell^2,\ T(x) = (\langle x, g_k\rangle)_k$
$T^*\colon \ell^2 \to H,\ T^*(y) = \sum_k y_k g_k$

Theorem 1: Frame Operator

The frame operator $S = T^*T$ satisfies:

(i) $S^* = S$.
(ii) S is positive definite.
(iii) S has a square root $S^{1/2}$.
(iv) S is invertible.

Theorem 2: Frame Bounds

The tightest possible frame bounds are given by the smallest and largest eigenvalues of S.

Theorem 3: Dual Frame

For a frame $(g_k)_k$ with frame bounds A and B, the set $(\tilde{g}_k)_k = (S^{-1}g_k)_k$ is frame with frame bounds $1/B$ and $1/A$, and with analysis operator $\tilde{T} = TS^{-1}$. We have $\tilde{T}^*T = T^*\tilde{T} = \mathrm{id}$.

Definition 1.3 (Tight Frame)**.** *A frame with frame bounds $A = B$ is called a tight frame.*

Theorem 4: Tight Frame

A frame is tight with frame bound A iff its frame operator is of the form $S = A\,\mathrm{id}$.

Theorem 5

For any frame $(g_k)_k$, the set $(S^{-1/2}g_k)_k$ is a tight frame with frame bound 1.

Theorem 6

A tight frame $(g_k)_k$ with frame bound $A = 1$ and $\|g_k\| = 1$ for all k is an ONB.

Definition 1.4. *A set $(g_k)_k$ is called complete, if*

$$\langle x, g_k\rangle = 0 \ \forall k \ \Rightarrow \ x = 0.$$

Definition 1.5. *A frame $(g_k)_k$ is called exact if for all n, the set $(g_k)_{k\neq n}$ is incomplete.*

Theorem 7: Exactness

Let $(g_k)_k$ be a frame and $(\tilde{g}_k)_k$ its canonical dual. Then, $(g_k)_k$ is exact iff $\langle g_j, \tilde{g}_k\rangle = \delta_{jk}$.

Theorem 8: Sampling theorem

Let $x \in L^2(\mathbb{R})$ be bandlimited to B, i.e. $\hat{x}(f) = 0$ if $|f| > B$. Then, for $1/T \geq 2B$,

$$x(t) = 2BT \sum_{k=-\infty}^{\infty} x(kT)\,\mathrm{sinc}(2B(t - kT)).$$

2 Uncertainty Relations

Definition 2.1 (Operator Norm)**.**

$$|||A||| = \max_{\|x\|=1} \|Ax\|.$$

Definition 2.2 (Frobenius Norm)**.**

$$\|A\| = \sqrt{\operatorname{tr}(AA^H)}.$$

Theorem 9

$$\frac{\|A\|}{\sqrt{\operatorname{rank} A}} \leq |||A||| \leq \|A\|.$$

Definition 2.3. *For unitary U, we set*

$$\Delta_{P,Q}(U) = \left|\left|\left| D_P U D_Q U^H \right|\right|\right|.$$

Theorem 10

$$\Delta_{P,Q}(U) = \max_{x\in \operatorname{im} UD_QU^H} \frac{\|D_P x\|}{\|x\|}.$$

Theorem 11

$$\frac{\sqrt{\operatorname{tr}(D_P U D_Q U^H)}}{\min\{|P|, |Q|\}} \leq \Delta_{P,Q}(U) \leq \sqrt{\operatorname{tr}(D_P U D_Q U^H)}.$$

Definition 2.4 (Coherence)**.** *For $A = (a_1 \ldots a_n) \in \mathbb{C}^{m\times n}$ with normalized columns, the coherence is defined as $\mu(A) = \max_{i\neq j} |a_i^H a_j|$.*

Theorem 12

$$\Delta_{P,Q}(U) \leq \sqrt{|P||Q|}\mu([I\ U]).$$

Definition 2.5. *Let $P \subseteq \{1, \ldots, m\}$ and $\varepsilon_P \in [0, 1]$. A vector $x \in \mathbb{C}^m$ is called ε_P-concentrated if $\|x - x_P\|_2 \leq \|\varepsilon_P\|x\|_2$.*

Theorem 13

Let $A, B \in \mathbb{C}^{m\times m}$ be unitary and $P, Q \subseteq \{1, \ldots, m\}$. Suppose there exists a nonzero ε_P-concentrated $p \in \mathbb{C}^m$ and a nonzero ε_Q-concentrated $q \in \mathbb{C}^m$ such that $Ap = Bq$. Then,

$$|P||Q| \geq \frac{[1 - \varepsilon_P - \varepsilon_Q]_+^2}{\mu([A\ B])^2}.$$

Theorem 14

Let $A, B \in \mathbb{C}^{m\times m}$ be unitary. If $Ap = Bq$ for nonzero $p, q \in \mathbb{C}^m$, then

$$\|p\|_0\|q\|_0 \geq \frac{1}{\mu([A\ B])^2}.$$

3 Compressive Sensing

Definition 3.1 (Spark)**.** *The spark of a matrix is denfined as the cardinality of the smallest set linearly dependent columns.*

Theorem 15

For a matrix $D \in \mathbb{C}^{m\times n}$, uniqueness of recovery of s-sparse vectors x from the observation $y = Dx$ is guaranteed if

$$\operatorname{spark}(D) > 2s.$$

Definition 3.2 (P0). $\arg\min \|x\|_0$ *subject to* $y = Dx$.

Theorem 16

$$1 + \frac{1}{\mu(D)} \le \operatorname{spark}(D)$$

Theorem 17

If

$$\|x_0\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(D)}\right),$$

then x_0 is the unique solution of (P0).

Definition 3.3 (P1). $\arg\min \|x'\|_1$ *subject to* $y = Dx'$.

Theorem 18

Let $y = Dx_0$ and assume x_0 has support set S. If

$$\max_{x\in\ker D\setminus\{0\}} \frac{\sum_{k\in S}|x_k|}{\sum_k |x_k|} < \frac{1}{2},$$

then x_0 is the unique solution of (P1).

Theorem 19

If,

$$\|x_0\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(D)}\right), \tag{1}$$

then

$$\max_{x\in\ker D\setminus\{0\}} \frac{\sum_{k\in S}|x_k|}{\sum_k |x_k|} < \frac{1}{2}.$$

Hence, (1) is a sufficient condition such that x_0 is the unique solution of (P1).

Theorem 20

If $D \in \mathbb{C}^{m\times n}$, then

$$\mu(D) \ge \sqrt{\frac{n-m}{m(n-1)}}.$$

Theorem 21

$$\|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{|\operatorname{supp} x|}\,\|x\|_2$$

4 Sampling Spectally Sparse Signals

Theorem 22

Consider a signal with spectral occupancy contained in I. To reconstruct the signal, we need

$$\lim_{r\to\infty} \inf_{t\in\mathbb{R}} \frac{|P \cap [t, t+r]|}{r} \ge |I|,$$

where $P = \{t_n\}_n$ denotes the sampling set.

5 Restricted Isometry Property

Definition 5.1. *For each $s = 1, \ldots, n$, the isometry constant δ_s of a matrix $\Phi \in \mathbb{C}^{m\times n}$ is the smallest numer such that*

$$(1-\delta_s)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\delta_s)\|x\|_2^2$$

for every s-sparse x.

Theorem 23

Let $y = \Phi x$. Assume that $\delta_{2s} < \sqrt{2} - 1$. Then, the solution x^* to

$$\arg\min \|x'\|_1 \quad \text{subject to } \Phi x' = y$$

satisfies

$$\|x^* - x\|_1 \le C_0 \|x - x_s\|_1$$

and

$$\|x^* - x\|_2 \le C_0 s^{-1/2} \|x - x_s\|_1$$

for some constant C_0. In particular, if x is s-sparse, recovery is exact.

Theorem 24

Let $y = \Phi x + n$. Assume that $\delta_{2s} < \sqrt{2} - 1$ and $\|n\|_2 \le \varepsilon$. Then, the solution x^* to

$$\arg\min \|x'\|_1 \quad \text{subject to } \|y - \Phi x'\| \le \varepsilon$$

satisfies

$$\|x^* - x\|_2 \le C_0 s^{-1/2} \|x - x_s\|_1 + C_1\varepsilon$$

for some constants $C_0, C1$.

Definition 5.2. *A matrix Φ is said to satisfy the restricted null-space property w.r.t S if*

$$\max_{x\in\ker\Phi\setminus\{0\}} \frac{\sum_{k\in S}|x_k|}{\sum_k |x_k|} < \frac{1}{2}.$$

Theorem 25

If the isometry constant of order $2s$ of Φ satisfies $\delta_{2s} < 1/3$, then Φ satisfies the restricted null-space property for any S with $|S| \le s$.

6 Johnson-Lindenstrauss Lemma

Theorem 26

Let $\varepsilon \in (0,1)$ and suppose

$$k \ge \frac{8}{\varepsilon^2 - \varepsilon^3} \log(2m).$$

Then, for every set $X \subset \mathbb{R}^n$ of m points, there exists a (linear) map $f\colon \mathbb{R}^n \to \mathbb{R}^k$ such that for all $x, x' \in X$ we have

$$(1-\varepsilon)\|x - x'\|^2 \le \|f(x) - f(x')\|^2 \le (1+\varepsilon)\|x - x'\|^2.$$

7 Approximation Theory

Consider a set $C \subseteq L^2(\Omega)$.

Definition 7.1. *Denote by*

$$\mathfrak{E}^l = \{E\colon C \to \{0,1\}^l\}$$

the set of binary encoders of length l and by

$$\mathfrak{D}^l = \{D\colon \{0,1\}^l \to C\}$$

the set of binary decoders of length l.

Definition 7.2. *The minimax code length $L(\varepsilon, C)$ for ε is*

$$L(\varepsilon, C) = \min\left\{l \in \mathbb{N} : \exists (E,D) \in \mathfrak{E}^l \times \mathfrak{D}^l : \sup_{f\in C} \|D(E(f)) - f\|_2 \le \varepsilon\right\}.$$

The optimal exponent $\gamma^*(C)$ is defined as

$$\gamma^*(C) = \sup\left\{\gamma \in \mathbb{R} : L(\varepsilon, C) \in O(\varepsilon^{-1/\gamma}),\ \varepsilon \to 0\right\}.$$

Definition 7.3. *A metric is a function $d : X \times X \to \mathbb{R}$ that satisfies*

1. $d(x,y) \geq 0$,
2. $d(x,y) = 0 \Leftrightarrow x = y$,
3. $d(x,y) = d(y,x)$,
4. $d(x,z) \leq d(x,y) + d(y,z)$.

Definition 7.4 (ε-covering)**.** *An ε-covering of a compact set C with respect to the metric d is a set $\{x_1, \ldots, x_n\} \subset C$ such that for any $x \in C$ there exists an x_i such that $d(x, x_i) \leq \varepsilon$.*

Definition 7.5 (Covering Number)**.** *The ε-covering number $N(\varepsilon, C, d)$ is the cardinality of the smallest ε-covering.*

Definition 7.6 (Metric Entropy)**.** *The metric entropy of C is defined as*

$$\log_2 N(\varepsilon, C, d).$$

Definition 7.7 (ε-packing)**.** *An ε-packing of a compact set C with respect to the metric d is a set $\{x_1, \ldots, x_n\} \subset C$ such that $d(x_i, x_j) > \varepsilon$ if $i \neq j$.*

Definition 7.8 (Packing Number)**.** *The ε-packing number $M(\varepsilon, C, d)$ is the cardinality of the largest ε-packing.*

Theorem 27

The packing and covering number are related according to

$$M(2\varepsilon, C, d) \leq N(\varepsilon, C, d) \leq M(\varepsilon, C, d).$$

8 Uniform Laws of Large Numbers

Theorem 28: Markov's Inequality

Let X be a random variable and assume $g : \mathbb{R} \to [0, \infty)$ is increasing. Then, for any c with $g(c) > 0$,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}g(x)}{g(c)}.$$

Theorem 29: Glivenko-Cantelli

For any distribution, the empirical CDF $\hat{F}_n$ satisfies

$$\|\hat{F}_n - F\|_\infty \to 0 \quad \text{a.s.}$$

Definition 8.1. *Let $\mathcal{F}$ be a set of integrable real-valued functions and let $\{X_i\}_{i=1}^n$ be a collection of i.i.d. samples from some distribution $\mathbb{P}$. Then we write*

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X) \right|.$$

Definition 8.2 (Glivenko-Cantelli Class)**.** *We say that $\mathcal{F}$ is a Glivenko-Cantelli class for $\mathbb{P}$ if $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}}$ converges to zero in probability as $n \to \infty$.*

Definition 8.3 (Rademacher Complexity)**.** *For any fixed collection $x_1^n = (x_1, \ldots, x_n)$, consider the set*

$$\mathcal{F}(x_1^n) = \{(f(x_1), \ldots, f(x_n) \mid f \in \mathcal{F})\}.$$

The empirical Rademacher complexity is defined as

$$\mathcal{R}(\mathcal{F}(x_1^n)/n) = \mathbb{E}_\varepsilon \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(x_i) \right| \right],$$

where $(\varepsilon_i)_{i=1}^n$ is a sequence of Rademacher random variables (uniform on $\{-1, +1\}$.)
The Rademacher complexity is defined as

$$\begin{aligned} \mathcal{R}_n(\mathcal{F}) &= \mathbb{E}_X \mathcal{R}(\mathcal{F}(x_1^n)/n) && (2) \\ &= \mathbb{E}_{\varepsilon, X} \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(x_i) \right| \right]. && (3) \end{aligned}$$

Theorem 30

For any b-uniformly bounded function class, i.e. $\|f\|_\infty \leq b$ for all $f \in \mathcal{F}$, any n and $\delta \geq 0$ we have

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \leq 2\mathcal{R}_n(\mathcal{F}) + \delta$$

with probability at least $1 - e^{-\frac{n\delta^2}{2b^2}}$. Consequently, if $\mathcal{R}_n(\mathcal{F}) = o(1)$, we have $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \to 0$ a.s.

Definition 8.4 (Polynomial Discrimination)**.** *A class $\mathcal{F}$ of functions has polynomial discrimination of order $\nu \geq 1$ if for each n and $x_1^n = (x_1, \ldots, x_n)$ the set $\mathcal{F}(x_1^n)$ has cardinality upper bounded according to*

$$|\mathcal{F}(x_1^n)| \leq (n+1)^\nu.$$

Theorem 31

Suppose that $\mathcal{F}$ has polynomial discrimination of order ν. Then, for all n and $x_1^n = (x_1, \ldots, x_n)$ we have

$$\mathcal{R}(\mathcal{F}(x_1^n)/n) \leq 4D(x_1^n) \sqrt{\frac{\nu \log(n+1)}{n}},$$

where $D(x_1^n) = \sup_{f \in \mathcal{F}} \sqrt{\frac{\sum_{i=1}^n f(x_i)^2}{n}}$.

Definition 8.5 (VC Dimension)**.** *Given a class $\mathcal{F}$ of binary-valued functions, we say that the set $x_1^n = (x_1, \ldots, x_n)$ is shattered by $\mathcal{F}$ if $|\mathcal{F}(x_1^n)| = 2n$. The VC dimension $\nu(\mathcal{F})$ is the largest integer n for which there is some collection $x_1^n = (x_1, \ldots, x_n)$ of n points that is shattered by $\mathcal{F}$.*

Theorem 32: Sauer-Shelah

Consider a set class S with $\nu(S) < \infty$. Then, for any collection of points $x_1^n = (x_1, \ldots, x_n)$ with $n \geq \nu(S)$ we have

$$|S(x_1^n)| \leq \sum_{i=0}^{\nu(S)} \binom{n}{i} \leq (n+1)^{\nu(S)}.$$